

Analisis Sentimen Masyarakat Terhadap Pinjaman Online di Twitter Menggunakan Algoritma Naïve Bayes Classifier dan K-Nearest Neighbor

Rival Afandi*, M Afdal, Rice Novita, Mustakim

Fakultas Sains dan Teknologi, Sistem Informasi, Universitas Islam Sultan Syarif Kasim Riau, Pekanbaru, Indonesia

Email: 12050314787@student.uin-suska.ac.id, m.afdal@uin-suska.ac.id, rice.novita@uin-suska.ac.id, mustakim@uin-suska.ac.id

Email Penulis Korespondensi: 12050314787@student.uin-suska.ac.id

Submitted: 07/06/2024; Accepted: 06/09/2024; Published: 07/06/2024

Abstrak—Perkembangan teknologi yang sangat pesat membawa pengaruh yang besar bagi manusia. Pengaruh dari perkembangan teknologi yang bisa kita rasakan yaitu di bidang keuangan. Salah satu yang cukup marak akhir-akhir ini yaitu pinjaman online. Pinjol atau pinjaman online adalah layanan pinjaman uang online yang cepat dan mudah melalui aplikasi atau situs web, dengan persetujuan dan pencairan cepat, namun sering kali memiliki bunga tinggi dan tenor pendek. Dalam Twitter, ulasan komentar dan informasi yang di pakai disimpan dalam bentuk teks. Salah satu proses pengambilan informasi text mining dalam katagori teks adalah Analisis Sentimen untuk melihat kecenderungan suatu sentimen atau pendapat apakah bersifat Positif, Negatif, atau Netral pada ulasan komentar pengguna aplikasi Pinjol. Pada hasil pengambilan data terdapat 600 data awal yakni 122 ulasan Positif, 432 ulasan Negatif dan 43 ulasan Netral. Kemudian proses klasifikasi sentimen menggunakan algoritma Naive Bayes dan K-NN menghasilkan akurasi, presisi dan recal yakni sebesar 68% ; 83% dan recal 74% pada algoritma Naive Bayes sedangkan hasil akurasi, presisi dan recal pada K-NN yakni 72% ; 74% dan recal 96% dengan percobaan menggunakan 80% data latih dan 20% data uji.

Kata Kunci: Analisis Sentimen; K-Nearest Neighbor; Naive Bayes; Pinjaman Online; Text Mining; Twitter

Abstract—The very rapid development of technology has had a big impact on humans. The influence of technological developments that we can feel is in the financial sector. One thing that is quite popular lately is online loans. Pinjol or online loan is a fast and easy online money lending service via an application or website, with fast approval and disbursement, but often has high interest and short tenors. On Twitter, review comments and information used are stored in text form. One of the processes for retrieving text mining information in the text category is Sentiment Analysis to see whether a sentiment or opinion tends to be Positive, Negative or Neutral in the reviews of Pinjol application user comments. In the data collection results there were 600 initial data, namely 122 Positive reviews, 432 Negative reviews and 43 Neutral reviews. Then the sentiment classification process using the Naive Bayes and K-NN algorithms produces accuracy, precision and recall of 68%; 83% and recall 74% on the Naive Bayes algorithm, while the results of accuracy, precision and recall on K-NN are 72%; 74% and recall 96% with experiments using 80% training data and 20% test data

Keywords: Sentiment Analysis; K-Nearest Neighbor; Naive Bayes; Online Loans; Text Mining; Twitter

1. PENDAHULUAN

Perkembangan teknologi yang sangat pesat membawa pengaruh yang besar bagi manusia. Pengaruh dari perkembangan teknologi yang bisa kita rasakan yaitu di bidang keuangan. Salah satu yang cukup marak akhir-akhir ini yaitu pinjaman online [1]. Yang mana pinjaman online merupakan perkembangan teknologi dari konsep peminjaman uang secara online, dan berakhir si peminjam akan mendapatkan bunga yang harus di bayar sesuai ketentuan yang telah di sepakati melalui sistem online [2]. Pada saat ini aplikasi dan situs pinjaman online sudah banyak beredar di kalangan Masyarakat dengan berbagai macam yang dengan mengusung tema dan motif yang sama [3].

Seiring dengan meningkatnya popularitas pinjaman online, analisis sentimen masyarakat terhadap layanan tersebut menjadi semakin penting [4]. Melalui pengawasan Otoritas Jasa Keuangan (OJK), Twitter menjadi salah satu platform utama untuk memantau pandangan publik terhadap pinjaman online [5]. Dengan banyaknya pendapat yang diberikan oleh masyarakat terhadap layanan jasa pinjaman online, maka masyarakat yang lain akan lebih berhati-hati lagi dalam memilih jasa pinjaman online. Guna mengantisipasi permasalahan yang akan muncul [6].

Masalah yang sering di temui ialah banyak pemberi pinjaman online yang belum memiliki regulasi yang memadai, sehingga potensi risiko bagi para peminjam menjadi lebih tinggi [7]. dengan hal itu dilakukannya analisis sentimen penilaian apa itu pinjaman online agar menjadi penting untuk mendapatkan pemahaman tentang pandangan dan opini masyarakat mengenai layanan ini [8]. Sentimen analisis atau opinion mining yang selalu mengacu pada pengolahan bahasa alami, komputasi linguistik dan text mining sekaligus bertujuan menganalisa pendapat, sentimen, evaluasi, sikap, penilaian dan emosi seseorang apakah pembicara atau penulis berkenan pada suatu topik, produk, layanan, organisasi, individu, ataupun kegiatan tertentu[9]. Oleh karna itu, Dalam proses analisis sentimen masyarakat terhadap Pinjaman online ini menggunakan teks mining dengan mengadopsi algoritma metode Naïve Bayes dan KNN. Algoritma tersebut merupakan metode analisis klasifikasi yang telah banyak digunakan. Algoritma memiliki keunggulan dan kelemahan dalam mengkasifikasikan data masing-masing, diantara Naïve Bayes Classifier dan Metode K-Nearest Neighbor algoritma mana yang paling baik dalam mengklasifikasikan data [10].

Penelitian dalam proses analisis sentimen masyarakat terhadap Pinjaman menggunakan teks mining dengan mengadopsi algoritma metode Naïve Bayes dan KNN. Algoritma tersebut merupakan metode analisis klasifikasi yang

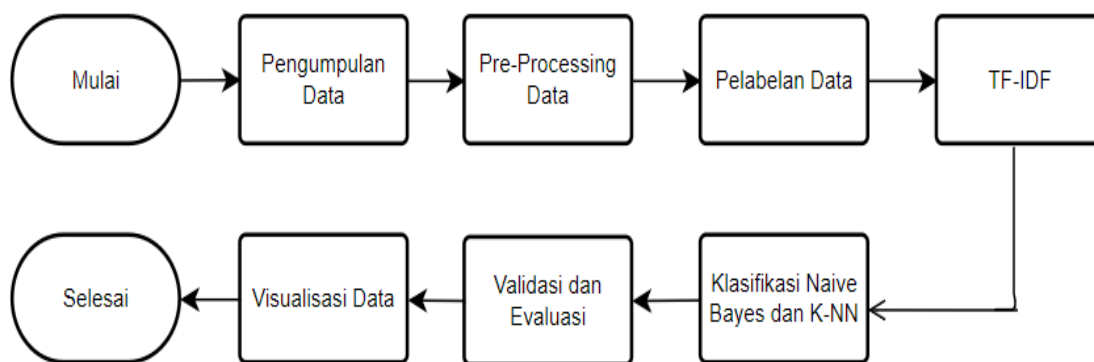
telah banyak digunakan. Algoritma memiliki keunggulan dan kelemahan dalam mengklasifikasikan data masing-masing, antara Naïve Bayes Classifier dan Metode K-Nearest Neighbor algoritma mana yang paling tepat dalam mengklasifikasikan data.

Penelitian ini menggunakan Metode *Naive Bayes* dan *K-Nearest Neighbord* (KNN) yang merupakan metode analisis sentimen yang cukup populer dan sering digunakan dalam penelitian-penelitian terkait. Salah satu penelitian terdahulu mengenai analisis sentimen yaitu menggunakan algoritma Naive Bayes terhadap pelanggan shopee. Jumlah data yang dipakai pada penelitian tersebut sebanyak 600 data. Hasil akurasi yang didapatkan dari pengujian dengan algoritma tersebut mencapai 97% [11]. Adapun penelitian yang di lakukan oleh yaitu Ikhsan Habib Kusuma dan Nuri Cahyono pada Analisis Sentimen Masyarakat Terhadap Penggunaan E-Commerce Menggunakan Algoritma K-Nearest Neighbor, yang mana menghasilkan 469 ulasan positif dan 426 ulasan negative. Hasil akurasi yang di dapat dari pengujian dengan algoritma K-Nearest Neighbor mencapai 82% [12]. Penelitian yang lainnya dalam pembahasan opini publik Covid-19 pada Twitter menggunakan metode Naïve Bayes dan KNN. Dengan menggunakan metode NB, Mendapatkan nilai akurasi sebesar 63.21% sedangkan nilai KNN sebesar 58.10% yang berarti metode NB paling akurat dibandingkan metode KNN [13]. . Penelitian tentang Sistem Analisis Sentimen pada ulasan produk dengan menggunakan total 1500 data dari femaledaily.com dan nilai akhir sebesar 77.78%, data tersebut kasifikasi menggunakan Naive Bayes [14].

Berdasarkan uraian dari permasalahan dan penelitian sebelumnya mengenai analisis sentiment, maka pada penelitian ini melakukan Analisis Sentiment Masyarakat Terhadap Aplikasi Pinjaman Online Pada Twitter Menggunakan Algoritma *Naïve Bayes Classifier* dan *K-Nearest Neighbor* untuk mengetahui bagaimana ulasan sentimen para pengguna terhadap aplikasi tersebut.

2. METODOLOGI PENELITIAN

Dalam penelitian ini, fokus utama ditujukan pada dua metode klasifikasi, yaitu Naïve Bayes dan K-Nearest Neighbor, meskipun ada berbagai metode klasifikasi lain yang tersedia. Proses untuk mendapatkan hasil yang optimal melibatkan beberapa tahapan, mulai dari pengumpulan dan pelabelan data, hingga proses preprocessing. Setelah data diproses, langkah selanjutnya adalah menerapkan algoritma Naïve Bayes dan KNN untuk klasifikasi. Gambar 1 merupakan tahapan-tahapan ini menjadi penting untuk memastikan kualitas dan akurasi hasil analisis.



Gambar 1. Alir Diagram Penelitian

2.1 Pengumpulan Data

Pada tahap awal penelitian, peneliti mengidentifikasi query tweet untuk mengumpulkan data. Query tweet yang dipilih untuk penelitian ini adalah istilah "Pinjaman Online", "Pinjol" dan "Depkolektor". Data yang digunakan berasal dari platform microblogging Twitter. Peneliti mengumpulkan data menggunakan metode yang disediakan oleh API Twitter dan menyimpannya dalam format csv

2.2 Pre-Processing

Pra-pemrosesan merujuk pada tahap persiapan data mentah sebelum diolah atau digunakan dalam model pembelajaran mesin. Tahap ini menjadi krusial dalam proses pembelajaran mesin karena kualitas dan sifat data dapat berdampak signifikan pada performa model. Proses preprocessing dilakukan untuk mengolah data mentah menjadi kumpulan data yang siap digunakan untuk analisis atau pemodelan. Untuk meningkatkan hasil preprocessing dari penelitian sebelumnya, penelitian ini menjalankan beberapa tahapan preprocessing, termasuklah:

- a. Case folding : Sebuah proses untuk mengedit teks dokumen ke dalam bentuk lower case



- b. Cleaning : Data yang digunakan perlu adanya proses pembersihan data seperti symbol tautan URL, angka. Data twitter sendiri tentunya banyak data kotor seperti tagar, angka, nama pengguna, URL dan teks retweet
- c. Tokenize : Proses data yang sebelumnya kalimat kemudian dipecah menjadi kata perkata
- d. Stopword Removal : Kata yang terdapat di dalam stoplist akan melalui tahap pembersihan
- e. Weighting : dalam proses ini untuk pembobotan kata dengan TF-IDF.

2.3 Pelabelan Data

Dataset yang dihasilkan dari keyword pinjol merupakan dataset yang belum berlabel atau masih berjenis unsupervised data. sehingga harus dilakukan pelabelan agar dataset dapat dipelajari. Pelabelan data dilakukan oleh 1 orang annotator. Annotator merupakan seorang yang bertanggung jawab untuk memberikan label (negatif, netral, dan positif) pada masing masing dataset.

2.4. TF-IDF

Tahap pembobotan dilakukan proses yaitu setiap kata diberi bobot dengan menggunakan Term Frequency dan Inverse Document Frequency (TF-IDF) [15]. Pada tahapan ini menggunakan Bahasa Pemrograman Python dengan Library Sk-Learn.

2.5. Klasifikasi Menggunakan Algoritma NBC dan K-NN

Pada penelitian ini dilakukan klasifikasi menggunakan dua algoritma yaitu NBC dan KNN

2.6 Naive Bayes

Naïve Bayes adalah algoritma yang terkenal dalam melakukan klasifikasi lebih tepatnya yaitu analisis sentiment. Pada algoritma tersebut didasarkan pada Teorema Bayes [12],[16]. Metode Naïve Bayes Classifier merupakan metode sederhana untuk klasifikasi probabilitas. Dalam metode ini, probabilitas dilakukan dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset. Secara analogi, penggunaan metode Naïve Bayes Classifier hanya membutuhkan sedikit data yang dimasukkan pada saat proses klasifikasi teks. Tujuan utama dari metode Naïve Bayes Classifier adalah untuk meminimalkan varians dalam dokumen.

2.7 K-Nearest Neighbor

Algoritma KNN merupakan sebuah algoritma pembelajaran yang dapat diawasi yang berguna untuk melakukan tugas klasifikasi serta regresi. Cara kerjanya adalah dengan mencari K titik data terdekat dari sampel yang diberikan, kemudian menggunakan label kelas atau nilai dari titik data tersebut untuk membuat prediksi terhadap label kelas atau nilai dari sampel tersebut. Dalam proses klasifikasi, KNN mengidentifikasi K titik data terdekat dalam set pelatihan dari sampel yang sedang diuji, berdasarkan jarak antara titik-titik tersebut seperti jarak Euclidean. Kemudian, algoritma ini menetapkan label kelas yang paling umum di antara K tetangga terdekat tersebut kepada sampel yang diuji.

2.8 Validasi dan Evaluasi

Pada tahap ini data yang telah dilakukan pembobotan akan divalidasi menggunakan metode Hold-Out. Dengan melakukan pengelompokan antara data uji dan data latih kemudian data akan dilakukan pemecahan data dengan membagi data menjadi data latih dan data uji.

2.9. Visualisasi dan Analisis

Visualisasi data adalah proses mengubah data mentah menjadi representasi grafis seperti grafik atau diagram untuk memudahkan pemahaman, analisis, dan komunikasi informasi. Ini membantu mengidentifikasi pola, tren, dan anomali dalam data, serta mendukung pengambilan keputusan berbasis data.

HASIL DAN PEMBAHASAN

3.1 Data awal

Pada penelitian analisis sentimen ini menggunakan data tweet hasil crawling menggunakan google colab dengan jumlah 600 data dengan rentan waktu April 2023 - Desember 2023 dengan kata kunci ‘pinjaman online’, ‘pinjol’ dan ‘depkolektor’. Dalam mengenai Pimjaman online atau Pinjol, dengan menggunakan metode Naive Bayes Classifier yang akan dibandingkan dengan Algoritma K-Nearest Neighbors (KNN). Tujuannya untuk mengetahui hasil perbandingan keberhasilan nilai accuracy, precision, dan recall. dengan beberapa tahap Preprocessing, Validasi dan Evaluasi pada proses preprocessing serta untuk memberikan label sentimen mengenai data tweet yang digolongkan menjadi label sentimen positif, negatif dan netral.

Tabel 1. Ulasan Komentar Pinjaman Oline

No	Tweet
----	-------



1	Posisi bank perekonomian rakyat (BPR) tergerus kehadiran bank umum dan layanan teknologi finansial (fintech) atau pinjaman online yang sama-sama menasar segmen mikro. #KoranTempo https://t.co/3qvjjN2GuS
2	mbak taylor Selamat pagi Kami Menawarkan Pinjaman ONLINE Tanpa Anggunan Bunga 2% tahun min 5juta-500juta Chat WA: 08122600960
3	Di luar ketentuan agama hati nurani saya bilang begini: Ini kalau saya pasang iklan ini ada yang tertarik terus kecanduan judi online lalu terlilit kasus pinjaman online saya salah satu orang yang membuat dia terjebak di situasi itu. So nope.
4	Komite Sekolah menghadirkan narasumber dari Bareskrim Polri AKBP Taat Resdianto (TN 10) yang memberi pencerahan terkait Bahaya Judi dan Pinjaman Online dan Liza Marielly Djaprie M.Psi. membahas aspek psikologi terkait Bahaya Judi dan Pinjaman Online. https://t.co/ssjrKV5u33
5	Unit Binmas Polsek Mergangsan sambang dan dialogis dengan Ibu Hanum warga Karang kajan Mergangsan terkait informasi pinjaman online Jum'at 17/11/2023. #PoldaDIY #polresjogja #polsekMergangsan https://t.co/ksX0abgkxn
...	..
600	Tahun 2024 nanti perusahaan financial technology (fintech) salah satunya pinjol akan menghadapi banyak tantangan setelah tahun 2023 dapat menunjukkan... https://t.co/FN2PGYVCg4

Dari hasil crawling data, data yang akan diambil hanyalah data yang ada pada table pada konten. Kemudian dilakukan proses Text Preprocessing.

3.2 Preprocessing

Data dari tweet atau komentar tentang Pinjaman Online dari Twitter sering tidak dapat langsung digunakan karena tidak terstruktur dan noise. Pre-processing data diperlukan untuk mempersiapkan dan membersihkan data agar lebih mudah digunakan [17][18]. Pre-processing dilakukan dengan Google Colab dan Python, mencakup pembersihan, pelipatan huruf, pemfilteran, dan stemming.

a. Tahap Case Folding

Case Folding adalah langkah pertama dalam persiapan dokumen. Dalam proses ini, semua huruf dalam dokumen teks diubah menjadi huruf yang lebih kecil. Hanya Huruf A sampai Z yang diterima [19]. Sehingga hasilnya seperti pada Tabel 2.

Tabel 2. Data Hasil Case Folding

No	Tweet
1	posisi bank perekonomian rakyat (bpr) tergerus kehadiran bank umum dan layanan teknologi finansial (fintech) atau pinjaman online yang sama-sama menasar segmen mikro. #korantempo https://t.co/3qvjjn2gus
2	mbak taylor selamat pagi kami menawarkan pinjaman online tanpa anggunan bunga 2% tahun min 5juta-500juta chat wa: 08122600960
3	di luar ketentuan agama hati nurani saya bilang begini: ini kalau saya pasang iklan ini ada yang tertarik terus kecanduan judi online lalu terlilit kasus pinjaman online saya salah satu orang yang membuat dia terjebak di situasi itu. so nope.
4	komite sekolah menghadirkan narasumber dari bareskrim polri akbp taat resdianto (tn 10) yang memberi pencerahan terkait bahaya judi dan pinjaman online dan liza marielly djaprie m.psi. membahas aspek psikologi terkait bahaya judi dan pinjaman online. https://t.co/ssjrkv5u33
5	unit binmas polsek mergangsan sambang dan dialogis dengan ibu hanum warga karang kajan mergangsan terkait informasi pinjaman online jum'at 17/11/2023. #poldadiy #polresjogja #polsekmergangan https://t.co/ksx0abgkxn
...	...
600	tahun 2024 nanti perusahaan financial technology (fintech) salah satunya pinjol akan menghadapi banyak tantangan setelah tahun 2023 dapat menunjukkan... https://t.co/fn2pgyvvcg4

Berikut merupakan sebuah hasil sebelum dan sesudah dilakukan case folding. Yang dimana karakter pada tabel sebelum case folding merupakan text asli. Kemudian menghasilkan karakter dalam string ke huruf besar-kecil. Kasus lipat dapat berguna dalam situasi di mana kasus teks tidak menyampaikan makna apapun, seperti dalam tugas pengolahan bahasa alami. Namun, itu mungkin tidak sesuai dalam situasi di mana kasus teks itu penting, seperti ketika menganalisis kata benda atau akronim yang tepat

b. Cleaning

Cleaning merupakan data yang diolah berupa teks akan dibersihkan terlebih dahulu untuk menghilangkan tanda baca, angka, simbol, link URL, username, dan kata duplikat di dalam teks [15]. Sehingga hasilnya jumlah datanya sekarang berjumlah seperti Tabel 3.

Tabel 3. Data Hasil Cleaning

No	Tweet
----	-------



1	posisi bank perekonomian rakyat tergerus kehadiran bank umum dan layanan teknologi finansial atau pinjaman online yang sama-sama menysasar segmen mikro
2	mbak taylor selamat pagi kami menawarkan pinjaman online tanpa anggungan bunga tahun min juta chat
3	di luar ketentuan agama hati nurani saya bilang begini: ini kalau saya pasang iklan ini ada yang tertarik terus kecanduan judi online lalu terlilit kasus pinjaman online saya salah satu orang yang membuat dia terjebak di situasi itu so nope
4	komite sekolah menghadirkan narasumber dari bareskrim polri akbp taat resdianto yang memberi pencerahan terkait bahaya judi dan pinjaman online dan liza marielly djaprie m psi membahas aspek psikologi terkait bahaya judi dan pinjaman online
5	unit binmas polsek mergangsan sambang dan dialogis dengan ibu hanum warga karang kajen mergangsan terkait informasi pinjaman online jum'at
...	...
600	tahun nanti perusahaan financial technology salah satunya pinjol akan menghadapi banyak tantangan setelah tahun dapat menunjukkan

c. Filtering

Setelah proses casefolding, dilakukan Filtering menggunakan stopword Tala yang merupakan stopwords untuk bahasa Indonesia. Artinya, hanya kata-kata yang menghasilkan kalimat positif, negatif, atau netral yang digunakan [20]. Tabel stopwords dapat dilihat pada Tabel 4.

Tabel 4. Kamus Stopword

No	Stopword	No	Stopword	No	Stopword	No	Stopword
1	ada	6	agaknya	11	akhiri	16	amatlah
2	adalah	7	agar	12	akhirnya	17	anda
3	adanya	8	akan	13	aku	18	andalah
4	adapun	9	akankah	14	akulah
5	agak	10	akhir	15	amat	1771	hapus

Setelah dilakukan Filtering menggunakan kamus stopwords, maka hasilnya akan seperti pada tabel 5.

Tabel 5. Data Hasil Filtering

No	Tweet
1	kompastv kok wakil yg protes cuma satu yg mengebu ngebu harusnya semua yg rapat wakil rakyat pada mengebu ngebu semua ntar orang lippobisa terkencing kencing dan takut lagi untuk berbuat manipulasitugas wakil rakyat harus lebih sanggar dari depkolektor dc
2	choymarkochoy totalpolitikcom pengamat yg slalu meleset hati2 alphard nya mgkin masih kredit nt depkolektor nagih lo hihhi
3	bayangin lu jadi depkolektor resiko kerjanya antara dapet duit digebukin atau meninggal
4	knpiharis pssi erickthohir nga ngaruh tuh urus aj depkolektor
5	partaisocmed ada pernah liat di daerah mana gtu ada larangan untuk depkolektor masuk daerah tsb coba serentak disemua tempat kaya gini
..	..
600	tahun 2024 nanti perusahaan financial technology fintech salah satunya pinjol akan menghadapi banyak tantangan setelah tahun 2023 dapat menunjukkan

d. Steaming

Setelah dilakukan Filtering, berikutnya dilakukan Stemming untuk proses mengambil kata dasar dengan membuang imbuhan kata di depan dan di belakang kata [21][22]. Sehingga hasilnya seperti pada Tabel 6.

Tabel 6. Data Hasil Steaming

No	Tweet
1	posisi bank ekonomi rakyat bpr gerus hadir bank layan teknologi finansial fintech pinjam online samasama sasar segmen mikro korantempo
2	mbak taylor selamat pagi tawar pinjam online anggun bunga 2 min 5juta500juta chat wa
3	tentu agama hati nurani bilang pasang iklan tarik candu judi online lilit pinjam online salah orang jebak situasi so nope
4	komite sekolah hadir narasumber bareskrim polri akbp taat resdianto tn 10 cerah kait bahaya judi pinjam online liza marielly djaprie mpsi bahas aspek psikologi kait bahaya judi pinjam online



5	unit binmas polsek mergangsan sambang dialogis hanum warga karang kajen mergangsan kait informasi pinjam online jumat poldadiy polresjogja polsekmergangan
...
600	usaha financial technology fintech salah satu pinjol hadap tantang

e. Pelabelan

Data Twitter yang ada pada Pinjaman Online merupakan daya yang tidak memiliki laber netral, positif dan negatif. Oleh karena itu, dibutuhkan metode khusus dalam menganalisis data *tweets* tersebut. Dalam penelitian ini, pelabelan data pinjaman online ditangani dengan manual oleh Dedi Febianto, S.Pd. M.Pd. Berikut merupakan hasil pelabelan *tweet* untuk pinjaman online seperti di Tabel 7.

Tabel 7. Potongan Hasil Pelabelan

No	Tweet	Label
1	posisi bank ekonomi rakyat bpr gerus hadir bank layan teknologi finansial fintech pinjam online samasama sasar segmen mikro korantempo	Netral
2	mbak taylor selamat pagi tawar pinjam online anggun bunga 2 min 5juta500juta chat wa	Netral
3	tentu agama hati nurani bilang pasang iklan tarik candu judi online lilit pinjam online salah orang jebak situasi so nope	Negatif
4	komite sekolah hadir narasumber bareskrim polri akbp taat resdianto tn 10 cerah kait bahaya judi pinjam online liza marielly djaprie mpsi bahas aspek psikologi kait bahaya judi pinjam online	Negatif
5	unit binmas polsek mergangsan sambang dialogis hanum warga karang kajen mergangsan kait informasi pinjam online jumat poldadiy polresjogja polsekmergangan	Positif
...	
600	usaha financial technology fintech salah satu pinjol hadap tantang	Netral

f. Term *Frequency-Inverse Document Frequency* (TF-IDF)

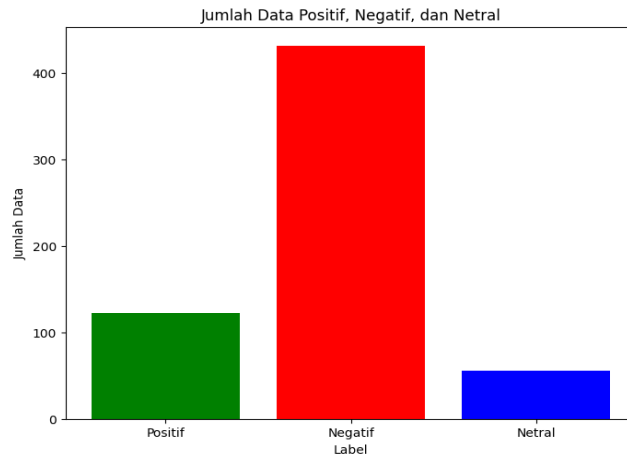
Data setelah pra-pemrosesan harus dalam bentuk numerik. Untuk ini, digunakan metode TF-IDF yang memberi bobot pada kata-kata dalam dokumen berdasarkan frekuensi kemunculan (tf) dan frekuensi terbalik dokumen (idf) [5]. Perhitungan TF-IDF dilakukan dengan Python menggunakan scikit-learn dan dibagi sesuai data latih dan uji. Hasilnya dapat dilihat pada Tabel 8.

Tabel 8. Potongan TF-IDF

No	pinalti	pinjam	pinjaman	bayar	pinjammeminjam	pinjeman	pinjol
1	0	0,084020886	0	0	0	0	0
2	0	0,107833058	0	0	0	0	0
3	0	0,0878424	0	0	0	0	0
4	0	0,128897954	0	0	0	0	0
5	0	0,074590608	0	0	0	0	0
6	0	0,081994127	0	0	0	0	0
7	0	0,073317058	0	0	0	0	0,06786973
8	0	0,106066944	0	0	0	0	1
..
600	0,07403	0	0	0	0	0,06853	0

3.3 Pembagian Data Latih dan Data Uji

Berdasarkan dengan metodologi penelitian yang telah dijelaskan, maka penelitian ini melakukan pembagian data latih (data training) dan data uji (data testing) yang diambil dari dataset dengan jumlah 600 data setelah melalui beberapa tahapan pre-processing dan tahap TF-IDF. Kemudian data tersebut akan digunakan pengujian menggunakan Algoritma NBC dan K-NN. Dari grafik yang terlihat pada gambar 2 terdapat 432 sentiment Negatif dan terdapat 122 sentiment Positif dan 43 Sentiment Netral..



Gambar 2. Grafik Jumlah Sentimen Positif dan Negatif

3.4 Evaluasi Model

3.4.1. Naive bayes

Pada penelitian ini data set dibagi dua bagian yaitu data uji dan data latih dengan rasio test dan train yaitu 20% data latih dan 80% data uji. Hasil uji akurasi pada metode naive bayes untuk klasifikasi yaitu sebesar 68% yang dapat dilihat pada table 9 sebagai berikut:

Tabel 9. Tabel Hasil Klasifikasi NBC

	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>	<i>Support</i>
<i>Negatif</i>	0.83	0.74	0.77	87
<i>Positif</i>	0.42	0.54	0.47	24
<i>Netral</i>	0.43	0.55	0.48	11
<i>Accuracy</i>			0.68	122

3.4.2 K-Nearest Neighbor (K-NN)

Selanjutnya menggunakan metode K-NN dengan hasil akurasi sebesar 72%, hasil dari metode K-NN dapat dilihat sebagai berikut:

Tabel 10. Hasil Akurasi Metode K-NN

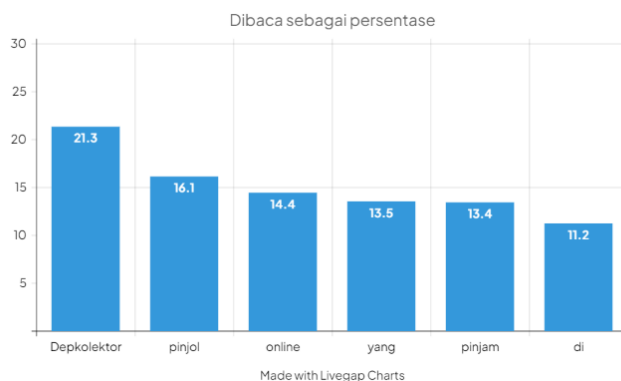
	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>	<i>Support</i>
<i>Negatif</i>	0.74	0.96	0.83	89
<i>Positif</i>	1.00	0.67	0.80	3
<i>Netral</i>	0.20	0.03	0.06	30
<i>Accuracy</i>			0.72	122

Tabel 9 dan 10 merupakan hasil dari pengujian confusion matrix. Dalam pengujian confusion matrik pada gambar diatas menghasilkan nilai akurasi, presisi dan recall. Berdasarkan dari analisis pengujian menggunakan metode NB dan KNN dari masing-masing metode dapat dirangkum hasilnya seperti tabel dibawah ini.

Tabel 11. Perbandingan Metode

Classifier	Accuracy	Precision	Recall
Naïve bayes	68%.	83%.	74%.
K-NN	72%.	74%.	96%.

Berdasarkan tabel 11 hasil perhitungan evaluasi yang telah dilakukan proses pengolahan data dengan metode NB dan KNN terlihat juga bahwa metode K-Nearest Neighbor terbaik untuk menyatakan nilai tingkat akurasi yang tinggi dengan nilai akurasi 72% diikuti oleh algoritma NB, dengan tingkat akurasi 68%.



Gambar 6. Kata Dengan Kemunculan Tertinggi

4. KESIMPULAN

Dua metode klasifikasi yang telah diterapkan pada analisis sentimen adalah Naive Bayes (NB) dan k-Nearest Neighbors (k-NN). Hasil penggunaan Naive Bayes menunjukkan akurasi sebesar 68%, sementara k-NN mencapai tingkat akurasi sebesar 72%. Namun, fokus pada sentimen negatif menunjukkan bahwa dari total 432 data negatif, NB berhasil mengklasifikasikan data 293 dengan benar, sedangkan k-NN berhasil mengklasifikasikan 313 data dengan benar. Begitu juga untuk sentimen positif, NB mengklasifikasikan 82 data dengan benar, sedangkan k-NN berhasil mengklasifikasi 87 data dengan benar dari total 122 data positif. Begitu juga untuk sentimen netral, NB mengklasifikasi 29 data dengan benar dari 43 data netral, sedangkan k-NN berhasil mengklasifikasikan 30 data dengan benar pada sentiment netral. Analisis ini memberikan gambaran yang lebih rinci tentang bagaimana kedua model berkinerja dalam mengklasifikasikan sentimen negatif dan positif secara spesifik. Meskipun k-NN memiliki akurasi yang lebih tinggi secara keseluruhan, NB menunjukkan kinerja yang lebih baik dalam mengklasifikasikan sentimen negatif, sementara k-NN lebih unggul dalam mengklasifikasikan sentimen positif. Dengan demikian, pemilihan model untuk analisis sentimen harus mempertimbangkan preferensi dan kebutuhan spesifik dalam menangani kelas sentimen yang berbeda. Analisis ini memberikan gambaran yang lebih rinci tentang bagaimana kedua model berkinerja dalam mengklasifikasikan sentimen negatif dan positif secara spesifik. Meskipun k-NN memiliki akurasi yang lebih tinggi secara keseluruhan, NB menunjukkan kinerja yang lebih baik dalam mengklasifikasikan sentimen negatif, sementara k-NN lebih unggul dalam mengklasifikasikan sentimen positif. Dengan demikian, pemilihan model untuk analisis sentimen harus mempertimbangkan preferensi dan kebutuhan spesifik dalam menangani kelas sentimen yang berbeda.

REFERENCES

- [1] A. F. I. Muhammad Imam Ghozali, Wibowo Harry Sugiharto, "Analisis Sentimen Pinjaman Online Di Media Sosial Twitter Menggunakan Metode Naive Bayes," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 33, no. 1, pp. 1–12, 2022, doi: 10.30865/klik.v3i6.936.
- [2] R. Wahyudi and G. Kusumawardhana, "Analisis Sentimen pada review Aplikasi Grab di Google Play Store Menggunakan Support Vector Machine," vol. 8, no. 2, pp. 200–207, 2021.
- [3] M. Diki Hendriyanto, A. A. Ridha, and U. Enri, "Analisis Sentimen Ulasan Aplikasi Mola Pada Google Play Store Menggunakan Algoritma Support Vector Machine Sentiment Analysis of Mola Application Reviews on Google Play Store Using Support Vector Machine Algorithm," *J. Inf. Technol. Comput. Sci.*, vol. 5, no. 1, pp. 1–7, 2022.
- [4] M. Ula, R. Zulhusna, R. Putra Fhonna, and A. Pratama, "Penerapan Model Klasifikasi K-Nearest Neighbor Dalam Pencarian Kesesuaian Pekerjaan," *Metik J.*, vol. 6, no. 1, pp. 18–23, 2022, doi: 10.47002/metik.v6i1.343.
- [5] D. A. Warraihan, I. Permana, M. Mustakim, and R. Novita, "Analisis Sentimen Pengguna Transportasi Online Maxim Pada Instagram Menggunakan Naive Bayes Classifier dan K-Nearest Neighbor," *J. Media Inform. Budidarma*, vol. 7, no. 3, pp. 1134–1143, 2023, doi: 10.30865/mib.v7i3.6336.
- [6] Artha Mahendra Putra and Rani Apriani, "Perlindungan Hukum Terhadap Konsumen Jasa Layanan Peminjaman Online Illegal," *The Juris*, vol. 6, no. 2, pp. 571–578, 2022, doi: 10.56301/juris.v6i2.630.
- [7] D. C. Arisona, G. N. A. Wibowo, S. Siswanto, and ..., "Klasifikasi Pesan Biasa, Operator, Spam, dan Debt Collector Menggunakan K-Nearest Neighbor. docx," *J. INSYPRO ...*, vol. 8, no. November, pp. 1–6, 2023, [Online]. Available: <https://journal.uin-alauddin.ac.id/index.php/insypro/article/view/41264%0Ahttps://journal.uin-alauddin.ac.id/index.php/insypro/article/download/41264/18442>
- [8] R. Wati *et al.*, "Analisis Sentimen Persepsi Publik Mengenai PPKM Pada Twitter Berbasis SVM Menggunakan Python," vol. 06, pp. 240–247, 2021.
- [9] L. Ardiani and H. Sujaini, "Implementasi Sentiment Analysis Tanggapan Masyarakat Terhadap Pembangunan di Kota Pontianak Implementation of Sentiment Analysis of Community Responses to Development in Pontianak City," vol. 8, no. 2, pp. 183–190, 2020, doi: 10.26418/justin.v8i2.36776.
- [10] H. Rhomadhona and J. Permadi, "Klasifikasi Berita Kriminal Menggunakan Naive Bayes Classifier (NBC) dengan Pengujian K-Fold Cross Validation," *J. Sains dan Inform.*, vol. 5, no. 2, pp. 108–117, 2019, doi: 10.34128/jsi.v5i2.177.
- [11] A. H. Ruger, M. Suyanto, and M. P. Kurniawan, "Sentimen Analisis Pelanggan Shopee di Twitter menggunakan Algoritma



- Naive Bayes,” vol. 1, no. 2, pp. 26–29, 2021.
- [12] I. Habib Kusuma and N. Cahyono, “Analisis Sentimen Masyarakat Terhadap Penggunaan E-Commerce Menggunakan Algoritma K-Nearest Neighbor,” *J. Inform. J. Pengemb. IT*, vol. 8, no. 3, pp. 302–307, 2023.
- [13] M. Syarifuddin, “Analisis Sentimen Opini Publik Mengenai Covid-19 Pada Twitter Menggunakan Metode Naïve Bayes Dan Knn,” *INTI Nusa Mandiri*, vol. 15, no. 1, pp. 23–28, 2020, doi: 10.33480/inti.v15i1.1347.
- [14] J. Edukasi, B. Gunawan, H. S. Pratiwi, and E. E. Pratama, “Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes,” vol. 4, no. 2, pp. 113–118, 2018.
- [15] A. V. Sudiantoro *et al.*, “Analisis Sentimen Twitter Menggunakan Text Mining Dengan,” vol. 10, no. 2, pp. 398–401, 2018.
- [16] Normah, B. Rifai, S. Vambudi, and R. Maulana, “Klasifikasi Penerima Bantuan Sosial Menggunakan Algoritma C 4.5,” *J. Tek. Komput. AMIK BSI*, vol. 8, no. 2, pp. 174–180, 2022, doi: 10.31294/jtk.v4i2.
- [17] Tania Puspa Rahayu Sanjaya, Ahmad Fauzi, and Anis Fitri Nur Masruriyah, “Analisis sentimen ulasan pada e-commerce shopee menggunakan algoritma naive bayes dan support vector machine,” *INFOTECH J. Inform. Teknol.*, vol. 4, no. 1, pp. 16–26, 2023, doi: 10.37373/infotech.v4i1.422.
- [18] M. P. Simatupang and D. P. Utomo, “Analisa Testimonial Dengan Menggunakan Algoritma Text Mining Dan Term Frequency- Inverse Document Frequence (Tf-Idf) Pada Toko Allmееart,” *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 3, no. 1, pp. 808–814, 2019, doi: 10.30865/komik.v3i1.1697.
- [19] D. Sebagai, S. Untuk, and M. Gelar, “NAIKNYA HARGA BBM PADA TWITTER MENGGUNAKAN ALGORITMA DECISION TREE-KNN-NAÏVE BAYES,” 2022.
- [20] H. Harianto and D. Rosiyadi, “Komparasi Algoritma C4.5, Naïve Bayes dan k-Nearest Neighbor Sebagai Sistem Pendukung Keputusan Menaikkan Jumlah Peserta Didik,” *J. Inform.*, vol. 7, no. 1, pp. 55–61, 2020, doi: 10.31311/ji.v7i1.7250.
- [21] A. D. Adhi Putra, “Analisis Sentimen pada Ulasan pengguna Aplikasi Bibit Dan Bareksa dengan Algoritma KNN,” *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 8, no. 2, pp. 636–646, 2021, doi: 10.35957/jatisi.v8i2.962.
- [22] J. C. Macuácuá, J. A. S. Centeno, and C. Amisse, “Data mining approach for dry bean seeds classification,” *Smart Agric. Technol.*, vol. 5, no. April, 2023, doi: 10.1016/j.atech.2023.100240.